DOCUMENT RESUME

ED 275 170 FL 016 074

AUTHOR Perkins, Kyle
TITLE A Regression Analysis of Direct and Indirect Measures of English as a Second Language Writing Compositions.
PUB DATE 84
NOTE 13p.; In: Practice and Problems in Language Testing. Papers from the International Symposium on Language Testing (7th, Colchester, England, 1984); see FL 016 066.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *English (Second Language); Higher Education; *Language Tests; Regression (Statistics); Research Methodology; *Scoring; Statistical Analysis; *Test Use; *Writing (Composition); Writing Evaluation; *Writing Skills

ABSTRACT

This study investigates the strength of relationships between direct measures (holistic scores for free writing samples) and indirect measures (the Test of Ability to Subordinate) of attained writing proficiency in English as a second language. Subjects were 67 undergraduate foreign language students enrolled in Basic Composition for Foreign Students at Southern Illinois University. Substantial magnitudes of correlation between the two measures were found. Stepwise multiple regression analysis indicated that both direct and indirect measures of writing accounted for substantial variance in a direct measure variable. Explanations of the remaining variance in the dependent variable are offered, limitations of the study are discussed, and suggestions are offered for future research using this paradigm. (Author/MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275170

A REGRESSION ANALYSIS OF DIRECT AND INDIRECT MEASURES OF ENGLISH AS A SECOND LANGUAGE WRITING COMPOSITIONS

KYLE PERKINS
(SOUTHERN ILLINOIS UNIVERSITY)

FL 016 074

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. Cuthane

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A REGRESSION ANALYSIS OF DIRECT AND INDIRECT MEASURES OF ENGLISH AS A SECOND LANGUAGE WRITING COMPOSITIONS

Kyle Perkins (Southern Illinois University)

Abstract

This study attempted to establish the strength of relationships between direct and indirect measures of attained English as a Second Language writing proficiency. The Test of Ability to Subordinate (TAS) was the indirect measure and holistic scores for free writing samples the direct measure. Substantial magnitudes of correlation were obtained between the TAS and the holistic evaluations. A stepwise multiple regression analysis indicated that both direct and indirect measures of writing accounted for substantial amounts of variance in a direct measure variable. Explanations are offered to account for the remaining variance in the dependent variable. Limitations of the study are discussed, and suggestions are offered for future research using this paradigm.

Introduction

There is considerable debate in the composition research community concerning the efficacy of direct and indirect tests of attained writing proficiency. Some researchers, like Michael and Shaeffer (1979) and Diederich (1974), advocate the use of multiple-choice tests of writing.

Michael and Shaeffer stated:

> A short multiple choice examination requiring knowledge of principles about grammar, usage, choice of words and phrases, and syntax such as the TSWE (Test of Standard Written English) or a subtest of the CSUC-EPT (California State University and Colleges English Placement Test) reflecting similar competencies provides a more accurate prediction of success in a basic English composition course emphasizing writing skills than does an essay examination requiring the actual demonstration of writing skills (p 132).

In a similar vein, Diederich noted:

> My final word of advice on such tests (objective tests) is not to despise them. Objective test items can easily, quickly and reliably test a student's knowledge of the rules and conventions of English. In writing, if a student is not sure that he knows how to use a certain construction he can change his sentence to avoid using it, but in an objective test, you can

> give him a sentence with that construction in it, and he has to decide whether it is correct or incorrect. Arguments over whether answering objective items 'is the same thing as' actual writing or speaking are futile. For that matter, writing one essay is not 'the same thing as' writing another essay, even on the same day; we have seen that even the most carefully determined grades on such essays rarely correlate higher than 0.70. Then the astonishing thing is that scores on a good objective test of English usage often correlate about 0.70 with averages of the two essay grades. It does not matter that they do not 'really' measure 'the same thing.' If students who are good at one tend to be good at the other, and vice versa, then it is a good indicator of proficiency in written English. Call it an editing test if you like, but I can promise you that students who do well on it tend to be good writers (pp 79-80).

But other researchers like Braddock, Lloyd-Jones and Schoer (1963) and Breland and Gaynor (1979) have criticized objective tests for their lack of validity and the impressions about the low valuation of writing which they convey.

Braddock, Lloyd-Jones and Schoer:

> The most serious charge against multiple-choice tests is their lack of validity. Not only do they require the examinee to perform the behavior being measured - he does no actual writing; but these tests also make no attempt to measure the 'large elements' of composition, even indirectly (p 42).

Breland and Gaynor:

> Indirect measures lack face validity and credibility among members of the English profession and educators generally, and they tend to deliver a message to students that writing is not important (p 127).

For further discussion on the correlation between direct and indirect test scores in composition, see Davis, Scriven and Thomas (1981); Hogan and Mishler (1980); Moss, Cole and Khampalikit (1982); and Stiggens (1982).

Research Questions

The research reported in this paper concerns (1) the strength of relationship between (a) direct and indirect measures of attained English as a second language (ESL) writing proficiency; (b) years of ESL study and direct and indirect measures of writing proficiency and (2) the amount of variance which a direct and an indirect measure and years of ESL study will account for in a direct measure of ESL writing.

More specifically, the following research questions were addressed:

1 Would a direct and an indirect measure of attained ESL writing proficiency correlate significantly?

2 Would years of ESL study and direct and indirect measures of ESL writing correlate significantly?

3 How much variance would years of ESL study, a direct and an indirect measure account for in a direct measure of ESL writing?

Procedures

Subjects This research was conducted in the Department of Linguistics, Southern Illinois University. The subjects were sixty seven undergraduate foreign language students enrolled in Basic Composition for Foreign Students. Native languages represented in the sample included Lao, Arabic, Farsi, Malay, Korean, Vietnamese, Thai, Chinese, Japanese, Igbo, Greek, Portuguese, Spanish, Norwegian and Gujerati.

Materials and Procedures At the beginning of a fifteen-week semester the Test of Ability to subordinate (TAS) was administered to the subject pool. During the same week the subjects wrote an expository composition which was graded holistically. During the testing procedure certain demographic data were collected, including the number of years each subject had studied English. During the final week of class the TAS was administered again and another expository composition was collected and graded holistically.

The TAS originates from a school of thought that holds that enhancement of syntactic fluency will also be reflected in improvement in overall writing ability. The TAS is considered, for this research, an indirect measure of writing, since it entails no text construction or actual composing. The following description of the TAS is found in the Teacher's Manual:

> The Test of Ability to Subordinate was designed primarily as a diagnostic instrument ot measure the ability of non-native speakers of English to control certain structures of subordination in writing. It was developed to fill the need for an easily scoreable, objective test that would measure students' ability to manipulate particular grammatical structures generally regarded as indicative of mature writing . . . the primary diagnostic purpose of the TAS is to identify which of nine structures of subordination students can manipulate and which they cannot, but a careful analysis of responses will highlight difficulties, if

any, with subject-verb agreement, plurals, tenses and articles. A teacher can use the information from the TAS to group students in terms of their strengths and weaknesses in subordination and to develop a writing program that will enable them to use these structures effectively as well as deal with other grammatical problems.

The essays were graded holistically by the author and a colleague. A five-point scale was employed: 5:excellent - 1:bad. Each paper was evaluated for form and content. To increase the variance the author summed all scores (rather than average the two evaluators' marks) for each paper. For example, say paper #33 received the following scores: Grader A, form 3; content 3; Grader B, form 4; content 5. The holistic score recorded for paper #33 was 15. The scoring was done independently and anonymously by each reader in an attempt to preclude any halo effect or contamination from affecting the results.

Analyses

The summary statistics and estimates of reliability for the writing measures are reported in Table 1. Both the TAS and holistic scoring exhibited acceptable coefficients of reliability for research of this nature.

TABLE 1

SUMMARY STATISTICS

Variable	Mean	S.D.	Type of reliability	Coefficient of reliability
L2 years	8.67	4.41		
Pre-TAS	31.85	7.41	KR-21	.85
Post-TAS	35.21	6.64	KR-21	.85
Pre-holistic	11.28	3.61	Interrater	.90
Post-holistic	11.25	3.58	Interrater	.92

The data generated for each variable were submitted to the Pearson product-moment correlation procedure. The correlation matrix is shown in Table 2.

TABLE 2

CORRELATION MATRIX

r_{xy}

		1	2	3	4	5
1.	L2 years	1.00				
2.	Pre-TAS	.48*	1.00			
3.	Post-TAS	.32*	.82*	1.00		
4.	Pre-holistic	.27	.75*	.69*	1.00	
5.	Post-holistic	.19	.56*	.51*	.55*	1.00

* $p < .01$, one-tailed test

Since the magnitude of any correlation coefficient is affected by the reliabilities of the measures of the X and Y variables, it is necessary to correct the correlation coefficients to hold reliability constant - the so-called correction for attenuation procedure. The disattenuated correlation matrix is presented in Table 3.

It can be seen in Table 3 that all correlations are significant at the .01 level but for L2 years, pre-holistic, and post-holistic. However, with large sample sizes such as this one (df=65) it is possible for comparatively small correlation coefficients to be found significant. Since coefficients above 0.50 are generally thought to be meaningful for research purposes, the correlations of 0.48 between L2 years and pre-TAS, and 0.32 between L2 years and post-TAS will be ignored, although they are significant at df65.

TABLE 3

CORRELATION MATRIX CORRECTED FOR ATTENUATION

r_{CA}

	1	2	3	4	5
1. L2 years	1.00				
2. Pre-TAS	.48*	1.00			
3. Post-TAS	.32*	.96*	1.00		
4. Pre-holistic	.27	.86*	.79*	1.00	
5. Post-holistic	.19	.64*	.58*	.60*	1.00

* $p < .01$, one-tailed test

A stepwise multiple regression analysis was conducted to determine how much variance the pre-TAS, post-TAS, pre-holistic and L2 years variables would account for in the post-holistic variable. The post-holistic scores were chosen as the criterion variable for the following reasons: (1) a direct writing sample at the end of a fifteen week semester of composition should be indicative of student writing proficiency and (2) like Braddock, Lloyd-Jones and Schoer, and Breland and Gaynor, the author prefers direct instead of indirect measures of writing. For ESL placement and proficiency testing the author prefers holistic evaluation of direct writing samples for the following reasons.

Holistic scoring has the highest construct validity of any composition evaluation scheme, when overall attained writing proficiency is the construct to be assessed. With holistic grading (1) the entire text is evaluated; (2) the grader can determine whether a thesis has been clearly stated, developed and supported or whether an issue has been clearly raised and sufficiently resolved; whether sufficient support and development (i.e., redundancy in the communication science sense) have been provided for the reader; whether the writer has tried to

accommodate the reader in terms of how much (how little) the writer thinks the reader knows about the subject (i.e., has the writer tried to accommodate the intended audience?); whether there is grammatical and lexical cohesion and overall coherence; whether one paragraph is 'hinged' smoothly to preceding and succeeding paragraphs through the accomplished use of transitional words, phrases and structure vocabulary; whether new information has been converted into given information that forms the base for predicating additional new information; and finally whether the piece 'makes sense'.

The results of the stepwise multiple regression are presented in Table 4.

TABLE 4

STEPWISE MULTIPLE REGRESSION

Criterion Variable: Post-holistic

Step	Source	R	R^2	R^2 INC	R^2 ADJ	DF	F INC	F TOT
1	Pre-TAS	.56	.314	.314	.303	1,65	29.75	29.75
2	Pre-holistic	.59	.352	.038	.332	1,64	3.75	17.38
3	L2 years	.60	.356	.004	.325	1,63	.39	11.61
4	Post-TAS	.60	.358	.002	.317	1,62	.19	8.64

Results and Discussion

As Table 3 indicates, the highest correlations obtained were 0.96 between pre- and post-TAS; 0.86 between pre-TAS and pre-holistic; 0.79 between post-TAS and pre-holistic; and 0.64 between pre-TAS and post-holistic. The largest coefficient represents a 92% overlap the two administrations of the TAS and could be interpreted as a coefficient of test-reliability. The three remaining coefficients which were cited could be viewed as

evidence that an indirect measure of writing, TAS, correlates substantially with a direct measure of writing, the holistic evaluations. The coefficient of 0.86 between pre-TAS and pre-holistic is correlational evidence of the TAS's concurrent validity with a 74% variance overlap. Pre-holistic and post-holistic correlated 0.79 - the 62% overlap being another correlational piece of evidence that a direct and indirect measure of writing can share substantial amounts of variance. And finally, there is a less substantial correlation of 0.64 which obtained between pre-TAS and post-holistic - a measure of the TAS's predictive validity. In sum, there seems to be evidence from the disattenuated correlation matrix that a direct and an indirect measure of ESL writing proficiency will share substantial magnitudes of covariance.

A more rigorous examination can be found in the stepwise regression analysis. The purpose of a multiple regression is to analyse the relationship between a dependent criterion variable (in this study, the post-holistic scores) and a set of independent, predictor variables (in this study, L2 years, pre-TAS, post-TAS and pre-holistic). For purposes of this research the linear dependence of post-holistic on L2 years, pre-TAS, post-TAS and pre-holistic is summarized and decomposed.

Table 4 illustrates the order in which the predictor variables entered the equation and the amount of variance accounted for by itself and cumulatively in linear combination with other predictor variables. Pre-TAS was the most robust predictor accounting for 31.40% of variance in the post-holistic variable. Pre-holistic accounted for an additional 3.80% of variance, the two variables together accounting for only 0.60% of the dependent variance (R^2_{INC} .004 + .002 = .006 x 100 = 0.60%).

At first blush it appears that the stepwise multiple-regression analysis establishes predictive validity for the TAS and the holistic evaluation and fails to establish concurrent validity for the TAS. These interpretations can be drawn by examining the F-statistics; however, the disappointing aspect of Table 4 is that the four predictor variables account for only 35.80% of the variance in the dependent variable, leaving 64.20% of the dependent variance unaccounted for by the predictor variables. It would be facetious to expect the predictor variables to account for 100% of the reliable variance in the dependent variable. On the other hand, the researcher believes that if the TAS were a really robust predictor variable, as well as the pre-holistic, 64.20% of the variance would not have been left unaccounted for. The point being made here is that for this kind of research, the magnitude of the r, R, and R^2 coefficients is more important than their levels of tabled value significance. Perhaps the findings can be explicated in the following section of the paper.

Limitations of the Study

There are at least two major limitations to this study. First, only one essay was evaluated for the pre-holistic scores and only one essay for the post-holistic scores. At least two essays should have been evaluated for each data collection for the following reason. Research has shown student writers vary significantly in their day-to-day performance in the composing process (Anderson 1960; Arthur 1979; Braddock-Lloyd-Jones and Schoer 1963; Kincaid 1953). The elicitation of multiple writing samples from the subject pool at various times on varying topics would have controlled for the fact that writing ability may vary with topic and time of day. Secondly, the different essays should have been elicited from different modes of discourse, because research has shown that writing students' ability varies across discourse forms (Arena 1975; Britton et al 1975; Kincaid 1953; Kinneary 1971; Lloyd-Jones 1977). Had these limitations been attended to, there might have been a significant difference between the pre-and post-holistic means (Table 1) and substantially more variance accounted for in the dependent variable (Table 4).

Although a multiple R or 0.60 (Table 4) is not shamefully low, the author would prefer to have more variance in the dependent variable accounted for by the predictor variables. There are five potential explanations as to why the multiple R was not larger: 1) the dependent or independent variables were not reliable; 2) the predictors were not robust; 3) there was a trait present which was not tapped by the measures; 4) there was redundancy present in the independent variables; 5) there was multicollinearity present in the independent variables. The first two explanations can be rejected out of hand, because the lowest reliability estimate for any variable was 0.85% and two independent variables did exhibit significant F-statistics in the regression equation. It would be impossible to empirically test claims three and four without a replication of this design with different instrumentation. The remaining factor, multicollinearity, is offered then as an explanation for the variance which was unaccounted for in the dependent variable. Multicollinearity is present when some or all of the independent variables are very highly intercorrelated. An inspection of Table 3 reveals this condition to be the case for this particular study. When the independent variables are very highly intercorrelated, the reliability of the relative importance indicated by the partial regression coefficients is reduced.

Conclusion

This study attempted to establish the strength of relationships between direct and indirect measures of attained ESL writing proficiency. The TAS was used as an indirect measure of writing and holistic evaluations of form and content were used as direct measures. Both the direct and indirect measures exhibited high estimates of reliability. Substantial magnitudes

of correlation were obtained between the two administrations of the TAS and the TAS and the holistic evaluations. A stepwise multiple regression analysis indicated that the pre-TAS and pre-holistic scores accounted for a significant amount of variance in the criterion variable, but all the predictor variables cumulatively accounted for only 35.80% of the variance in the dependent variable. Multicollinearity was suggested as the factor that accounts for lack of a larger multiple R. It was suggested that the lack of multiple writing samples from different discourse modes may pose limitations to the external validity of the study.

Such a study needs to be replicated to determine whether the results reported herein are valid. Other indirect measures such as the TSWE, the Cooperative English Test Effectiveness and Mechanics subtest, or the Revision and Editing Test (Matalene 1982) could be used as independent variables and holistic scores or primary trait scores, the criterion variables.

References

Anderson, C.C. 1960. The New STEP Essay Test as a Measure of Composition Ability. Educational and Psychological Measurement 20, pp 95-102

Arena, L. 1975. A Method for Improving the Writing Skills of Foreign Students in University-Level Expository English Composition Courses. In Burt, M.K. and Dulay, H.C. (eds) On TESOL '75: New Directions in Second Language Learning, Teaching, and Bilingual Education Washington D.C.: TESOL, pp 281-291

Arthur, B. 1975. Short-Term Changes in EFL Composition Skills. In Yorio, C.A., Perkins, K. and Schachter, J. (eds) On TESOL '79: The Learner in Focus Washington D.C.: TESOL, pp 330-342

Braddock, R., Lloyd-Jones, R. and Schoer, L. 1963. Research in Written Composition Urbana, IL: NCTE

Breland, H.M. and Gaynor, J.L. 1979. A Comparison of Direct and Indirect Assessments of Writing Skill. Journal of Educational Measurement 16, pp 119-128

Britton, J., Burgess, T., Martin, N., McLeod, A. and Rosen, H. 1975. The Development of Writing Abilities (11-18) London: Macmillan Education Ltd

Davis, B., Scriven, M. and Thomas, S. 1981. The Evaluation of Composition Instruction Pt. Reyes, CA: Edgepress

Diederich, P.B. 1974. Measuring Growth in English Urbana, IL: NCTE

Hogan, T. and Mishler, C. 1980. Relationships Between Essay Tests and Objective Tests of Language Skills for Elementary School Students. Journal of Educational Measurement 17, pp 219-227

Kincaid, G.L. 1953. Some Factors Affecting Variations in the Quality of Students' Writing Unpublished doctoral dissertation, Michigan State University

Kinneavy, J.L. 1971. A Theory of Discourse Englewood Cliffs, NJ: Prentice-Hall

Lloyd-Jones, R. 1977. Primary Trait Scoring. In Cooper, C.R. and Odell, L. (eds.) Evaluating Writing: Describing, Measuring, Judging Urbana, IL: NCTE, pp 33-66

Matalene, C. 1982. Objective Testing: Politics, Problems, Possibilities. College English 44, pp 368-381

Michael, W. and Schaeffer, P. 1979. A Comparison of the Validity of the Test of Standard Written English (TSWE) and of the California State University and Colleges English Placement Test (CSUC-EPT) in the Prediction of Grades in a Basic English Composition Course and of Overall Freshman-Year Grade Point Average. Educational and Psychological Measurement 39

Moss, P., Cole, N. and Khampalikit, C. 1982. A Comparison of Procedures to Assess Written Language Skills in Grade 10, 7, and 4. Journal of Educational Measurement 19, pp 37-48

Stiggins, R. 1982. An Analysis of Direct and Indirect Methods of Writing Assessment. Research in the Teaching of English 16, pp 57-69